23. März 2000

# PCT

## REQUEST

The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty.

For receiving Office use only

International Application No.

International Filing Date

Name of receiving Office and "PCT International Application"

Applicant's or agent's file reference *(if desired) (12 characters maximum)* 19741P WO

**Box No. I TITLE OF INVENTION**

Detection, cloning and sequencing of polypeptides which drive the subcellular localization of proteins

**Box No. II APPLICANT**

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

Europäisches Laboratorium für
Molekularbiologie (EMBL)
Meyerhofstraße 1
69117 Heidelberg
DE

☐ This person is also inventor.

Telephone No.

Facsimile No.

Teleprinter No.

State *(that is, country)* of nationality: DE

State *(that is, country)* of residence: DE

This person is applicant for the purposes of: ☐ all designated States ☒ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

**Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)**

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

GONZALEZ Caystano
Erwin-Rohde Straße 22
69120 Heidelberg
DE

This person is:

☐ applicant only

☒ applicant and inventor

☐ inventor only *(If this check-box is marked, do not fill in below.)*

State *(that is, country)* of nationality: ES

State *(that is, country)* of residence: DE

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☒ the United States of America only ☐ the States indicated in the Supplemental Box

☒ Further applicants and/or (further) inventors are indicated on a continuation sheet.

**Box No. IV AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE**

The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as: ☐ agent ☐ common representative

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

Weickmann H., Weickmann F.A., Huber B.,
Liska H., Prechtel J., Böhm B., Weiß W.,
Tiesmeyer J., Herzog M., Ruttensperger B., Jordan V.
Kopernikusstraße 9, 81679 München

Telephone No. 089/ 455 63-0

Facsimile No. 089/ 455 63-999

Teleprinter No. 522 621 wepat d

☐ **Address for correspondence:** Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.

Continuation of Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

*If none of the following sub-boxes is used, this sheet should not be included in the request.*

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

BEJARANO Luis
Boxbergring 107
69126 Heidelberg
DE

This person is:
- ☐ applicant only
- ☒ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

State *(that is, country)* of nationality: ES

State *(that is, country)* of residence: DE

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☒ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:
- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:
- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:
- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

☐ Further applicants and/or (further) inventors are indicated on another continuation sheet.

Box No.V DESIGNATION OF STATES

The following designations are hereby made under Rule 4.9(a) *(mark the applicable check-boxes; at least one must be marked)*:

Regional Patent

☐ AP ARIPO Patent: GH Ghana, GM Gambia, KE Kenya, LS Lesotho, MW Malawi, SD Sudan, SL Sierra Leone, SZ Swaziland, TZ United Republic of Tanzania, UG Uganda, ZW Zimbabwe, and any other State which is a Contracting State of the Harare Protocol and of the PCT

☐ EA Eurasian Patent: AM Armenia, AZ Azerbaijan, BY Belarus, KG Kyrgyzstan, KZ Kazakhstan, MD Republic of Moldova, RU Russian Federation, TJ Tajikistan, TM Turkmenistan, and any other State which is a Contracting State of the Eurasian Patent Convention and of the PCT

☒ EP European Patent: AT Austria, BE Belgium, CH and LI Switzerland and Liechtenstein, CY Cyprus, DE Germany, DK Denmark, ES Spain, FI Finland, FR France, GB United Kingdom, GR Greece, IE Ireland, IT Italy, LU Luxembourg, MC Monaco, NL Netherlands, PT Portugal, SE Sweden, and any other State which is a Contracting State of the European Patent Convention and of the PCT

☐ OA OAPI Patent: BF Burkina Faso, BJ Benin, CF Central African Republic, CG Congo, CI Côte d'Ivoire, CM Cameroon, GA Gabon, GN Guinea, GW Guinea-Bissau, ML Mali, MR Mauritania, NE Niger, SN Senegal, TD Chad, TG Togo, and any other State which is a member State of OAPI and a Contracting State of the PCT *(if other kind of protection or treatment desired, specify on dotted line)* .....

National Patent *(if other kind of protection or treatment desired, specify on dotted line)*:

☐ AE United Arab Emirates
☐ AL Albania
☐ AM Armenia
☐ AT Austria
☒ AU Australia
☐ AZ Azerbaijan
☐ BA Bosnia and Herzegovina
☐ BB Barbados
☐ BG Bulgaria
☐ BR Brazil
☐ BY Belarus
☒ CA Canada
☐ CH and LI Switzerland and Liechtenstein
☒ CN China
☐ CR Costa Rica
☐ CU Cuba
☐ CZ Czech Republic
☐ DE Germany
☐ DK Denmark
☐ DM Dominica
☐ EE Estonia
☐ ES Spain
☐ FI Finland
☐ GB United Kingdom
☐ GD Grenada
☐ GE Georgia
☐ GH Ghana
☐ GM Gambia
☐ HR Croatia
☐ HU Hungary
☐ ID Indonesia
☐ IL Israel
☐ IN India
☐ IS Iceland
☒ JP Japan
☐ KE Kenya
☐ KG Kyrgyzstan
☐ KP Democratic People's Republic of Korea
☐ KR Republic of Korea
☐ KZ Kazakhstan
☐ LC Saint Lucia
☐ LK Sri Lanka
☐ LR Liberia
☐ LS Lesotho
☐ LT Lithuania
☐ LU Luxembourg
☐ LV Latvia
☐ MA Morocco
☐ MD Republic of Moldova
☐ MG Madagascar
☐ MK The former Yugoslav Republic of Macedonia
☐ MN Mongolia
☐ MW Malawi
☐ MX Mexico
☐ NO Norway
☐ NZ New Zealand
☐ PL Poland
☐ PT Portugal
☐ RO Romania
☐ RU Russian Federation
☐ SD Sudan
☐ SE Sweden
☐ SG Singapore
☐ SI Slovenia
☐ SK Slovakia
☐ SL Sierra Leone
☐ TJ Tajikistan
☐ TM Turkmenistan
☐ TR Turkey
☐ TT Trinidad and Tobago
☐ TZ United Republic of Tanzania
☐ UA Ukraine
☐ UG Uganda
☒ US United States of America
☐ UZ Uzbekistan
☐ VN Viet Nam
☐ YU Yugoslavia
☐ ZA South Africa
☐ ZW Zimbabwe

Check-boxes reserved for designating States which have become party to the PCT after issuance of this sheet:

☐ DZ Algeria
☐ AG Antigua and Barbuda

**Precautionary Designation Statement:** In addition to the designations made above, the applicant also makes under Rule 4.9(b) all other designations which would be permitted under the PCT except any designation(s) indicated in the Supplemental Box as being excluded from the scope of this statement. The applicant declares that those additional designations are subject to confirmation and that any designation which is not confirmed before the expiration of 15 months from the priority date is to be regarded as withdrawn by the applicant at the expiration of that time limit. *(Confirmation (including fees) must reach the receiving Office within the 15-month time limit.)*

**Box No. VI PRIORITY CLAIM** ☐ Further priority claims are indicated in the Supplemental Box.

| Filing date of earlier application (day/month/year) | Number of earlier application | Where earlier application is: national application: country | regional application:* regional Office | international application: receiving Office |
|---|---|---|---|---|
| item (1) Oct 18, 1999 | 99 120 622.8 | | EP | |
| item (2) March 23, 1999 | 99 105 846.2 | | EP | |
| item (3) | | | | |

☐ The receiving Office is requested to prepare and transmit to the International Bureau a certified copy of the earlier application(s) *(only if the earlier application was filed with the Office which for the purposes of the present international application is the receiving Office)* identified above as item(s): _____

** Where the earlier application is an ARIPO application, it is mandatory to indicate in the Supplemental Box at least one country party to the Paris Convention for the Protection of Industrial Property for which that earlier application was filed (Rule 4.10(b)(ii)). See Supplemental Box.*

**Box No. VII INTERNATIONAL SEARCHING AUTHORITY**

**Choice of International Searching Authority (ISA)** *(if two or more International Searching Authorities are competent to carry out the international search, indicate the Authority chosen; the two-letter code may be used)*: ISA /

**Request to use results of earlier search; reference to that search** *(if an earlier search has been carried out by or requested from the International Searching Authority)*: Date *(day/month/year)* Number Country *(or regional Office)*

**Box No. VIII CHECK LIST; LANGUAGE OF FILING**

This international application contains the following **number of sheets**:

| | |
|---|---|
| request | : 4 |
| description (excluding sequence listing part) | : 18 |
| claims | : 4 |
| abstract | : 1 |
| drawings | : 8 |
| sequence listing part of description | : 2 |
| **Total number of sheets** | : 37 |

This international application is **accompanied** by the item(s) marked below:

1. ☒ fee calculation sheet
2. ☐ separate signed power of attorney
3. ☐ copy of general power of attorney; reference number, if any:
4. ☐ statement explaining lack of signature
5. ☐ priority document(s) identified in Box No. VI as item(s):
6. ☐ translation of international application into (language):
7. ☐ separate indications concerning deposited microorganism or other biological material
8. ☐ nucleotide and/or amino acid sequence listing in computer readable form
9. ☐ other *(specify)*:

**Figure of the drawings** which should accompany the abstract:

**Language of filing** of the international application: English

**Box No. IX SIGNATURE OF APPLICANT OR AGENT**

*Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the request).*

23. März 2000

B. Huber

**For receiving Office use only**

1. Date of actual receipt of the purported international application:
2. Drawings: ☐ received: ☐ not received:
3. Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application:
4. Date of timely receipt of the required corrections under PCT Article 11(2):
5. International Searching Authority (if two or more are competent): ISA /
6. ☐ Transmittal of search copy delayed until search fee is paid.

**For International Bureau use only**

Date of receipt of the record copy by the International Bureau:

# PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

## PCT

### NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Commissioner
US Department of Commerce
United States Patent and Trademark Office, PCT
2011 South Clark Place Room 524
Arlington, VA 22202
ETATS-UNIS D'AMERIQUE
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

| | |
|---|---|
| **Date of mailing** (day/month/year)<br>26 October 2000 (26.10.00) | |
| **International application No.**<br>PCT/EP00/02607 | **Applicant's or agent's file reference**<br>19741P WO |
| **International filing date** (day/month/year)<br>23 March 2000 (23.03.00) | **Priority date** (day/month/year)<br>23 March 1999 (23.03.99) |
| **Applicant**<br>GONZALEZ, Caystano et al | |

1. The designated Office is hereby notified of its election made:

[X] in the demand filed with the International Preliminary Examining Authority on:

28 September 2000 (28.09.00)

[ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was

[ ] was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

| | |
|---|---|
| **The International Bureau of WIPO**<br>**34, chemin des Colombettes**<br>**1211 Geneva 20, Switzerland** | Authorized officer<br>R. E. Stoffel |
| Facsimile No.: (41-22) 740.14.35 | Telephone No.: (41-22) 338.83.38 |

PATENT COOPERATION TREATY

REC'D 01 MAY 2001
WIPO PCT

# PCT

# INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

| Applicant's or agent's file reference | FOR FURTHER ACTION | See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416) |
|---|---|---|
| 19741P WO | | |
| International application No. | International filing date *(day/month/year)* | Priority date *(day/month/year)* |
| PCT/EP00/02607 | 23/03/2000 | 23/03/1999 |
| International Patent Classification (IPC) or national classification and IPC<br>C12N15/10 | | |
| Applicant<br>EUROPÄISCHES LABORATORIUM FÜR MOLEKULARBIOLOGIE.. | | |

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 5 sheets, including this cover sheet.

   ☒ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of 1 sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☐ Priority
   - III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☐ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   - VI ☐ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☒ Certain observations on the international application

| Date of submission of the demand | Date of completion of this report |
|---|---|
| 28/09/2000 | 26.04.2001 |
| Name and mailing address of the international preliminary examining authority:<br>European Patent Office - Gitschiner Str. 103<br>D-10958 Berlin<br>Tel. +49 30 25901 - 0<br>Fax: +49 30 25901 - 840 | Authorized officer<br>Korsner, S-E<br>Telephone No. +49 30 25901 329 |

# INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/EP00/02607

## I. Basis of the report

1. This report has been drawn on the basis of *(substitute sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments (Rules 70.16 and 70.17).)*:
**Description, pages:**

| | | | | | |
|---|---|---|---|---|---|
| 1-18 | as published | | | | |

**Claims, No.:**

| | | | | | |
|---|---|---|---|---|---|
| 1-15 | as published | | | | |
| 16-20 | as received on | 06/03/2001 | with letter of | 06/03/2001 | |

**Drawings, sheets:**

| | | | | | |
|---|---|---|---|---|---|
| 1/8-8/8 | as published | | | | |

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

   These elements were available or furnished to this Authority in the following language: , which is:

   ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).
   ☐ the language of publication of the international application (under Rule 48.3(b)).
   ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

   ☐ contained in the international application in written form.
   ☐ filed together with the international application in computer readable form.
   ☐ furnished subsequently to this Authority in written form.
   ☐ furnished subsequently to this Authority in computer readable form.
   ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
   ☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. The amendments have resulted in the cancellation of:

- ☐ the description, pages:
- ☐ the claims, Nos.:
- ☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

   *(Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)*

6. Additional observations, if necessary:

**V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

1. Statement

| | | | |
|---|---|---|---|
| Novelty (N) | Yes: | Claims | 1-20 |
| | No: | Claims | |
| Inventive step (IS) | Yes: | Claims | 1-20 |
| | No: | Claims | |
| Industrial applicability (IA) | Yes: | Claims | 1-20 |
| | No: | Claims | |

2. Citations and explanations
   **see separate sheet**

**VIII. Certain observations on the international application**

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:
**see separate sheet**

## V. Reasoned statement

Initial note:
The two priority documents have been checked during the international phase and it is considered that they cover the claimed subject-matter; the cited P,X-document ("publication of the invention") of the 17th of November, 1999, is therefore not relevant.

Novelty (Article 33(2) PCT)

The claimed processes (Claims 1-9, 10-11 and 12-15) for i.a. detecting, identifying and sequencing the intended polypeptides as well as the use (Claims 16-20) of the thereby identified polypeptides are novel over the prior art.

Inventive step (Article 33(3) PCT)

Although it was known in the prior art to use GFP for identifying intracellular localization of proteins (see Proc. Natl. Acad. Sci. USA; 1996, pages 15146-15151), the present concept relates to (i) the identification of peptides and their subcellular localizations in combination with (ii) the use of such peptides as carriers for directing (=the term "drive" of the claims) proteins to a predetermined subcellular location, for instance.

This concept is inventive over the prior art.

## VIII. Certain observations

The following observations are intended for consideration in a later phase:

Claims:

1.

Since the polypeptide and "part thereof" of Claims 16-20 are defined by the method of preparation/detection only (but not by any formula), the use cannot be fully searched and examined.

The possibility of accidental overlap with other (fusion) proteins cannot be excluded.

This would apply mutatis mutandis to any corresponding (but unidentified) nucleic acid sequences and vectors.

See also Claim 10 concerning the wording "...identification and/or production of a protein..."; should the identified protein or "part thereof" happen to be known, its production may not be patentable.

2.

The document Proc. Natl. Acad. Sci. USA as discussed above should preferably be identified in the Description as relevant background art.

- - - - -

19741PWO/BBcl 06. März 2001

New claims 16-20

16. Use of a polypeptide or part thereof, which drives the subcellular localization of a protein containing such polypeptide or part thereof, and which is detected and/or cloned according to any one of claims 1 to 15 in a vector for the expression of a desired protein wherein the vector contains a specific site into which a DNA encoding said desired protein can be inserted,
**characterized in that** the vector further comprises a DNA sequence encoding a polypeptide or a part thereof which drives the subcellular localization of a protein containing such polypeptide or part thereof, which DNA sequence is positioned in such a way that a fusion protein of desired protein and polypeptide or part thereof is encoded.

17. Use according to claim 16,
**characterized in that** the vector is a eucaryotic vector.

18. Use according to claim 16 or 17,
**characterized in that** the vector further comprises a reporter gene positioned in such a way that a fusion protein of desired protein and polypeptide or part thereof and reporter gene product is encoded.

19. Use according to claim 18,
**characterized in that** the reporter gene product is visually detectable.

20. Use according to anyoner of claims 16 to 19,
**characterized in that** the vector further contains sequences encoding proteolytic cleavage sites between one or more of the constituents of the fusion protein.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/10, 15/12, 15/62, C07K 14/435

A1

(11) International Publication Number: WO 00/56875

(43) International Publication Date: 28 September 2000 (28.09.00)

(21) International Application Number: PCT/EP00/02607

(22) International Filing Date: 23 March 2000 (23.03.00)

(30) Priority Data:

| | | |
|---|---|---|
| 99105846.2 | 23 March 1999 (23.03.99) | EP |
| 99120622.8 | 18 October 1999 (18.10.99) | EP |

(71) Applicant *(for all designated States except US)*: EUROPÄISCHES LABORATORIUM FÜR MOLEKULARBIOLOGIE (EMBL) [DE/DE]; Meyerhofstrasse 1, D–69117 Heidelberg (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: GONZALEZ, Caystano [ES/DE]; Erwin–Rohde Strasse 22, D–69120 Heidelberg (DE). BEJARANO, Luis [ES/DE]; Boxbergring 107, D–69126 Heidelberg (DE).

(74) Agents: WEICKMANN, H. et al.; Kopernikusstrasse 9, D–81679 München (DE).

(81) Designated States: AU, CA, CN, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

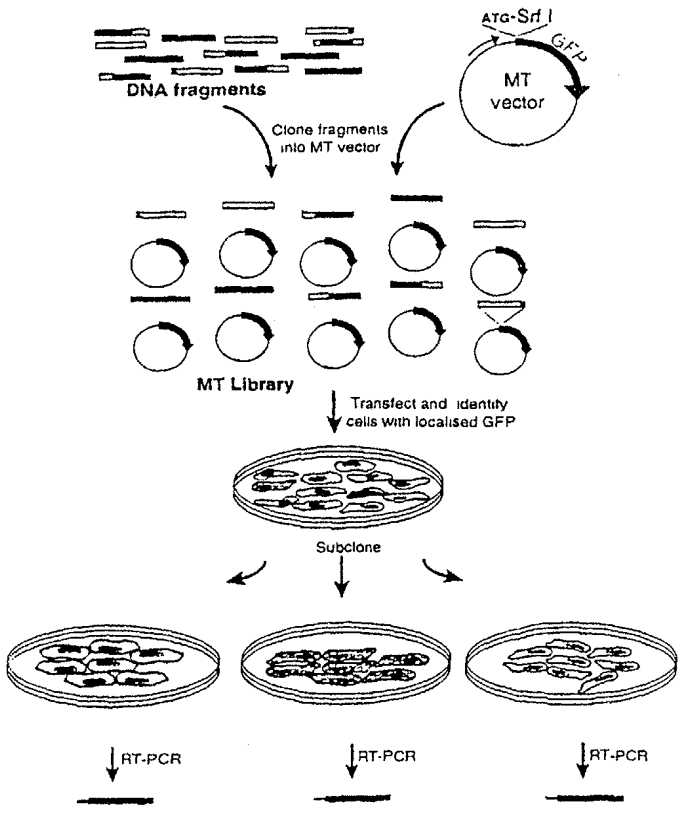
*With international search report.*

*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: DETECTION, CLONING AND SEQUENCING OF POLYPEPTIDES WHICH DRIVE THE SUBCELLULAR LOCALIZATION OF PROTEINS

(57) Abstract

The present invention concerns a process for the detection, cloning and/or sequencing of polypeptides or parts thereof, which drive the subcellular localization of a protein containing such polypeptide or part thereof, characterized in that the process comprises the following steps: (a) constructing an expression library of random nucleic acids ligated to a reporter gene and contained in a vector molecule, (b) transfecting a plurality of host cells with the library, (c) screening for the subcellular localization of the expression product of the nucleic acid in the host cells via detection of a signal produced by the reporter gene, (d) cloning such cells where the reporter gene signal is detected in a certain subcellular localization, and (e) cloning and optionally sequencing the nucleic acid insert which encodes the polypeptide or part thereof. Polypeptides, driving the intracellular localization can be used to construct fusion proteins with predetermined intracellular localization.



# Detection, cloning and sequencing of polypeptides which drive the subcellular localization of proteins

## Specification

The present invention relates to a process for the detection, cloning and/or sequencing of polypeptides or parts thereof, which drive the subcellular localization of a protein containing such polypeptide or part thereof, a process for the identification and/or production of a protein that is localized in a given subcellular localization, and a process for directing the subcellular localization of a nucleic acid expression product.

One of the most conspicuous features of the eukaryotic cell is its high degree of compartmentalization. Chromatin, nuclear matrix, nuclear membrane, Golgi apparatus, endoplasmatic reticulum, the endo- and exocytic compartments, the actin and microtubule cytoskeletons, mitochondria, the centrosome and the cell membrane are just some of the major subcellular organelles/compartments which have been defined by standard cytological analysis. Moreover, most of these compartments can be further subdivided into well differentiated regions or structures having a cytological and molecular identity of their own, thus resulting in the large number of subcellular domains which characterize the eukaryotic cell.

The physiological relevance of such compartmentalization is paramount. Every major cellular activity can be assigned to one or more well defined subcellular compartments. As a matter of fact, the intricate regulatory networks which operate within eukaryotic cells greatly rely on the differential subcellular localization of the molecules involved. The close relationship between subcellular localization and function is such that in most instances determining the preferential subcellular localization of a protein provides one of the best clues as to its putative function.

The molecular basis of specific subcellular localizations is not yet well understood. In some cases the localized protein contains a functional domain which drives its targeting either directly or through interaction with other previously localized members of the target structure. Well known examples of this first case are the nuclear localization signals which are recognized by the nuclear pore complex and translocated into the nucleus or the combination of a polybasic domain and a C-terminal CAAX motif, which leads to the post-translational modification of the protein and its membrane targeting. The second case includes kinesins and MAPs the cytoskeletal, spindle or centrosomal localization of which is achieved by virtue of their interaction with microtubuli. Finally, it is also conceivable that other mechanisms, e.g. different rates of import, export and degradation, could result in steady-states which may account for the preferential subcellular localization of proteins which do not contain any bona fide subcellular localization signal of their own.

In view of the diversity of mechanisms which may account for the subcellular localization of a given protein, it was an object of the present invention to provide a possibility to detect signals which either directly or indirectly drive subcellular localization. Such signals are usually polypeptides or parts thereof that are present in proteins. The knowledge of such polypeptides can be useful for a plurality of assays or applications.

The object underlying the invention was accomplished by the provided process for the detection, cloning and/or sequencing of polypeptides or parts thereof, which drive the subcellular localization of a protein containing such polypeptide, wherein said process comprises:

(a) constructing an expression library of random nucleic acids ligated to a reporter gene and contained in a vector molecule,

(b) transfecting a plurality of host cells with the library,

(c) screening for the subcellular localization of the expression product of the nucleic acid in the host cells via detection of a signal produced by the reporter gene,

(d) cloning such cells where the reporter gene signal is detected in a certain subcellular localization of interest, and

(e) cloning and optionally sequencing the nucleic acid insert which encodes the polypeptide or part thereof.

The above process allows for the detection of polypeptides or parts thereof, which drive the targeting of a protein to a particular subcellular location or structure, said process being completely independent of the function, organization and length of the respective protein containing such polypeptide. The process according to the invention can also be used to detect polypeptides or parts thereof that relocate intracellulary under specific conditions, like stress for example following heat shock or infection with a pathogen, all of which result in a dramatic rearrangement of the architecture of the cell (Cudmore et al., Trends Microbiol. (1997) 5, 142-147). Finally also polypeptides or parts thereof can be detected or cleaved according to the method of the invention that mediate the retention of proteins at specific organelle structures or loci.

Depending on the length of the random nucleic acids the process of the invention also allows the detection of complete or nearly complete proteins, that due to the presence of a polypeptide driving the localization are transfered to a certain intracellular location after expression. According to the invention random nucleic acids are produced from the genome of an organism. Either genomic DNA or cDNA may be used to generate the random nucleic acids that are ligated to a reporter molecule and inserted in vector molecules to construct the expression library of feature (a). Random nucleic acid molecules that are produced by subjecting the DNA of interest to restriction cleavage are only detected in the location of interest if they contain at least such a portion of said polypeptide which ensures that the

subcellular localization driving function is retained. Other constructs might also lead to expression of a fusion protein displaying the reporter gene signal albeit not in the localization of interest.

Out-of-frame insertions are generally not expressed in the process according to the invention.

It is irrelevant whether the random nucleic acid and reporter gene are ligated before being inserted in a vector molecule together, or whether a vector containing the reporter gene (lateron also termed GET (GFP-epitope trap) vector) is constructed into which a random nucleic acid can easily be ligated in an appropriate location next to the reporter gene. It is generally preferred to produce a fusion gene product, with the reporter gene located at the C-terminus. If the reporter gene is located at the N-terminal end, it will always be expressed thereby increasing background. Nevertheless, this might be the only way to detect some proteins which will not get localized if the fusion is made in the other direction. Theoretically, it is also possible that the reporter gene product is located in between nucleic acid expression products, as long as both the reporter gene and/or the polypeptide or part thereof retain their function.

The thus obtained vector molecules containing the random nucleic acid and reporter gene are transfected into a plurality of host cells which are then subjected to conditions allowing for the expression of the vector insert, whereupon screening for the subcellular localization of the expression product of the nucleic acid in the host cells takes place via detection of a signal produced by the reporter gene. Cells which show the reporter gene signal in a subcellular localization of interest are selected and subcloned, whereupon the DNA sequence insert can be cloned and optionally sequenced, said nucleic acid insert encoding the polypeptide or part thereof.

A schematic presentation of the process according to the invention is shown in Fig. 1. Fig. 1 shows the GFP-epitope trap approach. Random DNA fragments cloned into the GET vector will produce fusion proteins between the polypeptides encoded by these inserts and GFP. The GET vector contains a high efficiency cloning site immediate after the initial ATG of the GFP which shifts this codon out of frame with the rest of the coding sequence. Thus, GFP can only be expressed from vectors carrying an insert that restores the reading frame. Upon transfection with the GET library, a fraction of the cells can be observed to express GFP which in some cases will be localised in a particular compartment or organelle. These cells can be cloned and the inserts that code for the relevant subcellular localisation signal can be isolated by RT-PCR.

In a preferred embodiment of the present invention, a cDNA is used to create the random nucleic acids, or an expression library made of a cDNA is used.

In a further preferred embodiment of the invention, a library, either genomic or cDNA, from a mammalian organism or yeast or C. elegans or C. laevis is used. This preferred embodiment is not intended to limit the invention, since DNA from any organism may be used to detect specific polypeptide or parts thereof which drive the subcellular localization in the respective organism.

In one embodiment of the invention, a homologous system of nucleic acid for the creation of the expression library and host cell for the transfection is used. The homologous expression system is meant to identify a system where host cells are used that belong to the same species from which the nucleic acid was obtained.

In another embodiment of the invention, a heterologous system of nucleic acid library and cells for the transfection is used. In such a case, the host

cell does not belong to the same species as the nucleic acid that is to be expressed therein.

An example for a heterologous system would be the use of a Drosophila DNA for the generation of the expression library and of mammalian or yeast cells as host cells to be transfected with the vector molecule.

For the process according to the invention, standard procedures can be used which are known to the man in the art. Cloning of cells can be done either manually, picking up the cells and replating them for as many times as required to isolate one clone, or by serial dilution. Also, a fax sorter may be used which separates individual cells expressing a specific marker.

In a preferred embodiment of the invention, a reporter gene leading to a visually detectable signal upon expression is used.

Although principally other reporter genes are suitable, too, the visually detectable expression product is most easily detectable. Especially preferred reporter genes are genes coding for GFP (green fluorescent protein), or GFP derivatives like for example BFP (blue fluorescent protein), luciferase, YFP (yellow fluorescent protein), or CFP (cyan fluorescent protein). These derivatives are described in Pepperkok et al., Current Biology 9: 269 - 272 (1999) and references quoted therein.

The process according to the invention makes it possible to establish a system allowing to screen a huge number of nucleic acid molecules for the presence of a sequence encoding a polypeptide or part thereof capable of driving the subcellular localization of a protein containing such polypeptide. The process according to the invention has also the advantage that it can preferably be used in higher eukaryotes.

Therefore, a further subject of the present invention is a process for identifying and/or producing a protein that is localized in a given subcellular location, such process comprising the above process according to the invention as well as the cloning of a nucleic acid coding for a polypeptide epitope driving the localization in this given subcellular localization, and the use of said cloned nucleic acid to detect longer DNA sequences coding for a protein containing such polypeptide epitope. This detection can be conducted by standard molecular biology techniques, for example by hybridizing the nucleic acid to a genomic or cDNA library from a certain species and detecting homologous sequences encoding a protein, by RT-PCR or by comparing the obtained nucleic acid sequence with databases containing a huge number of DNA sequences. Such databases contain sequences coding for known proteins as well as sequences which are postulated to be coding for a protein, which, however, has either not been identified yet or the function of which is still unknown. The process according to the invention, therefore, is also useful as a high throughput method of determination of the subcellular localisation of the fast growing number of sequences which are being generated or detected by ongoing genome projects.

As soon as, by the above process for the identification of a protein containing the polypeptide, the respective nucleic acid coding for such protein is obtained or the sequence thereof is known, said nucleic acid can be expressed in an expression system, producing said protein containing a polypeptide or used in any other way including formation of mutants etc.

Another application of the process according to the invention is the identification of the proteins that are differentially sorted in differentiating cells, like for example cells that are induced to polarize or primary cultures of differentiating neurons (Dotti and Simmons, Cell (1990), 62: 63-72).

A further interesting use of the present process will also be the cloning of interacting partners of a given protein by transfecting cells which contain the protein labelled with a fluorochrome that produces FRET (Cubitt, et al., TIBS (1995) 20: 448-455) with the library's reporter.

Finally, systematic screenings with a library produced according to the invention could be used to identify new domains within known organelles and compartments.

A still further subject of the present invention is a process for directing the subcellular localization of a nucleic acid expression product. Said process comprises detecting a polypeptide or part thereof driving the localization of a protein containing such polypeptide according to the above process of the invention, and obtaining the nucleotide sequence encoding such polypeptide or part thereof, wherein further the nucleic acid coding for the polypeptide or part thereof is fused to a nucleic acid coding for a protein to be expressed, and the fusion product is expressed.

In a preferred embodiment of the present invention, the nucleic acid for the polypeptide or part thereof driving the localization and a reporter gene are fused with the nucleic acid coding for a protein to be expressed. By such fusion of a polypeptide and a reporter gene with a nucleic acid, the actual expression of the protein to be expressed at the localization of interest can be monitored.

For this purpose, it is preferred that a reporter gene is used the expression product of which is visually detectable.

It is a further preferred but in no way obligatory embodiment of the invention that the fusion product of the protein to be expressed and polypeptide or part thereof and/or reporter gene contains a proteolytic cleavage site between the protein to be expressed and the polypeptide

and/or the reporter gene product. According to this preferred embodiment, it is possible to obtain the pure protein to be expressed in a given localization by cleaving off the part directing the localization and optionally also the part enabling the monitoring of the expression and localization.

To this end, it might be useful to also express a corresponding proteolytic enzyme and direct it to the same subcellular localization by means of the process of the invention.

A further subject matter of the present invention is a vector for the expression of a desired protein wherein the vector contains a specific site into which a DNA encoding said desired protein can be inserted, said vector being characterized by further comprising a DNA sequence encoding a polypeptide or a part thereof which drives the subcellular localisation of a protein containing such polypeptide or part thereof, which DNA sequence is positioned in such a way that a fusion protein of desired protein and polypeptide or part thereof is encoded.

According to the invention any vector which is suitable for gene expression in an envisaged expression system can be employed. In a preferred embodiment of the invention, the vector is a eucaryotic vector and the envisaged expression system a eucaryotic system. The specific site into which a DNA encoding said desire protein can be inserted preferably is a restriction site that allows an in frame expression of the DNAs encoding the desired protein and encoding the polypeptide or part thereof. The site can also be a polylinker containing several restriction sites.

As described above for the process of directing the subcellular localization of a nucleic acid expression product, also for the vector according to the invention it is preferable that the vector further contains a reporter gene in such a manner, that upon expression of the desired protein and the polypeptide driving the localization or the part thereof also a reporter gene

product is expressed. This reporter gene product can either be expressed in form of a fusion protein with the other two components, or it can be separately expressed as a separate fusion with polypeptide driving the localization or part thereof.

In all of these vector constructs encoding fusion proteins of polypeptide driving the localization and/or reporter gene it is further preferable that DNA sequences encoding proteolytic cleavage sites are present is such a position that after expression the components can be separated from each other, thus facilitating purification of the desired protein. In this connection a further vector may contain a gene coding for the proteolytic enzyme also in such a manner that it is connected with a polypeptide driving the intracellular localization or part thereof. It is also feasable that the same vector coding for a fusion protein also contains the DNA sequences necessary to encode a fusion which results in a proteolytic enzyme being expressed and localized to the same cellular compartment as the other fusion protein, containing the desired protein.

It is also possible to use the process according to the invention for the detection, cloning and/or sequencing of polypeptides or parts thereof, which drive the subcellular localization of a protein containing such polypeptide or part thereof, for establishing a cell line or a collection of cell lines which are tranformed with a vector according to the invention. Such cell lines may show a reporter gene product at different locations. Such cell lines or such collection of cell lines is a further subject of the present invention as well as a kit containing a vector or a cell line according to the invention and which is useful for the expression of a desired protein in a desired localisation of a host cell.

The following examples along with the accompanying figures are intended to further elucidate the invention:

Fig. 1 shows a presentation of the general process of the invention.

Fig. 2 A - D shows a schematic representation of the process steps performed in Example 1.

Fig. 3 shows examples of subcellular localization of a reporter protein.

Fig. 4 shows patterns of GFP localisation generated by transfection with a GET library. Low (A) and high (B to I) magnification views of HEK 293 cells were counterstained for DNA using propidium iodine (red). B) mitochromosomes (arrows); F) the mitotic spindle. We have not determined yet the subcellular localisation of GFP in the cells shown in panels 2 G, H and I. Scale bar = 15 $\mu$m.

Fig. 5 shows sequencing the inserts that target GFP localisation. A) The GFP fusion in clone 02/11#22 shows a strong nucleolar localisation with a faint homogeneous nuclear background. B) The insert from this clone contains a well defined bipartite NLS (red) and meets the consensus of a nucleolar localisation signal. C) In clone 09/07#18 GFP colocalised with the ER as shown by counterstanding with an antibody against $\alpha$-calnexin (not shown). D) The insert from this cell line encodes a peptide of 35 amino acids that contains a predicted trans-membrane motif (PMSIFQLIYFLLFLFLGVIC). This sequence does not have a match in the sequence databases. Scale bar = 15 $\mu$m.

**Example 1**

Creating the Td2 fragment in EGFP-N1

The vector pEGFP-N1 (CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303-4230, USA) was modified by PCR using the following primers:

Oligo A (5'-CATGTTGGCGGCCGCGGTACCGTCGA-3') (SEQ.ID.NO.1)
Oligo B (5'-GCCCGGGCGTGAGCAAGGGCGAG-3'). (SEQ.ID.NO.2)

Oligo A contains an ATG with a good Kozak and a SrfI site. The ATG is out of frame with the GFP coding sequence.

PCR was carried out with Expand High Fidelity PCR System (Boehringer Mannheim GmbH, Sandhofer Strasse 116, D-68305 Mannheim, Germany) as indicated in the following protocol:

| Step | Temperature | Time |
|---|---|---|
| 1 | 94°C | 3 min |
| 2 | 42°C | 1 min |
| 3 | 68°C | 4 min (slope 13°C/sg) |
| 4 | 94°C | 1 min |
| 5 | 42°C | 1 min |
| 6 | 68°C | 4 min (goes to 4, 7 cycles) |
| 7 | 68°C | 7 min |
| 8 | 4°C | pause |

The PCR product was purified using the PCR Quiaquick Purification System (QIAGEN GmbH, Max-Valmer-Strasse 4, 40724 Hilden, Germany) and ligated with Rapid Ligation Kit (Boehringer Mannheim GmbH, supra). The ligated vector was used to transfect Epicurian Coli XL1-Blue (Stratagene, 11011 North Torrey Pines Road, La Jolla, CA 92097) by heat shock. After transfection, the cells were plated out in LB-Agar supplemented with 30 $\mu$g/ml kanamycin and incubated for 16 hours at 37°C. The DNA of some of the resulting colonies was isolated by minipreps and analyzed with restriction enzymes to confirm that it corresponded to the expected modified vector.

From this modified vector, the Td2 fragment was purified and subcloned into the NotI site of vectors of the pQE30 series (previously modified to contain this site). As expected, only Td2 fragments cloned into the pQE30 NotI resulted in cells expressing GFP. After this check, the Td2 fragment was cloned back into the NotI site of pEGFP-N1NotI. This is a derivative of pEGFP-N1 into which a NotI site was introduced in the polylinker. The final vector is pEGFPTd2.

**Modification of the pQE30 series**

These were modified introducing between the Bam HI and KpnI sites an adapter containing a NotI site. The adapter was made by annealing the following oligos:

Not 1-1b (5'-GATCGCGGCCGCGTAC-3') (SEQ.ID.NO.3)

Not 1-8 (5'-GCGGCCGC-3'). (SEQ.ID.NO.4)

Fig. 2 A - D shows the procedures of Example 1 schematically.

**Example 2**

**Construction of the "epitope-trap" library**

Drosophila gDNA was cut to completion with AluI and HaeIII (Boehringer Mannheim, supra), purified with QIAEXII Gel Extraction Kit, run in agarose gel for further size selection, purified again with QIAEXII Gel Extraction Kit and cloned into the SrfI of pEGFPTd2. Ligated DNA was used to transform E. coli XL1-Blue MR (Stratagene, supra). A small fraction of the cells was plated in LB medium containing 30 $\mu$g/ml kanamycin. The resulting clones were isolated and their DNA was purified and analyzed to determine the size of the average insert. The results were about 420,000 clones, approx. 6,700 (1.6%) of which had no insert, the estimated average length of the inserts being around 490 bp.

The remaining cells were plated out over a sterile nylon filter laid on a 24x24 plate containing LB supplemented with kanamycin, incubated at 30°C for 24 hours, replicated into another filter and reincubated for 4 hours at 37°C. The DNA was then obtained using Plasmid Maxi kit by QIAGEN.

Cells were transfected with a library prepared as described above. Ten hours after transfection, cells were fixed with methanol, and observed under the microscope. The reporter protein used in this example (GFP) can be observed as a bright white. In each of the examples shown in Fig. 2, the reporter is localized in different components within the cells. Observations were made with a Leica TCS confocal microscope system (Leica, Germany).

**Example 3**

In a typical transfection experiment with HEK293 cells and the MmcDNA-GET library, about 50% of the cells express GFP of which 20% display a distinct localisation of this reporter. Around 8 to ten hours after transfection, some cells start to express GFP and the first localisation patterns are recognisable (Figure 4A). Figure 4B to I shows some of the GFP localisation patterns that we observed. Panel 4B shows GFP specficially localised in the mitochondria as confirmed by counterstaining with the mitochondria-specific marker mitotracker (not shown). In the cell shown in Figure 4C GFP displays is fairly uniform in the cytoplasm, but is significantly concentrated in a small area near the nucleus that corresponds to the centrosome (arrow) as revealed by counterstaining with a human, autoimmune anti-centrosome antibody (not shown). Panels 4D, E and F show mitotic cells from different GFP expressing lines. GFP can be seen to localise in the cytokinesis furrow (arrow; Figure 4D), the chromosomes (Figure 4E) and the mitotic spindle (Figure 4F). GFP does not appear to be localised during interphase in these two cells. We have not yet determined the precise subcellular localisation of GFP in the cells shown in Panels 4G, H and I.

These are a few examples of the patterns of GFP localisation that we have observed. Using GET we have been able to identify cells with GFP localised in every major organelle and compartment. These observations illustrate the power of GET to identify specific molecular associations to organelles and compartments. To demonstrate the use of GET to identify proteins sequences that carry targeting signals we have cloned and sequenced the DNA inserts from some of these cells. As expected, we have found sequences that correspond to known proteins and contain targeting signals which are consistent with the observed localisation of the GFP fusion. One of these is clone 02/11#22, (Figure 5A, B). The GFP fusion in this cell line shows a distinct nucleolar localisation with a weak nuclear background. The insert from this line is identical to a fragment that spans between amino acids 62 and 131 of the mouse homologue of the HTLV-I tax responsive element binding protein TAXREB107 (Nacken et al., Biochim Biophys Acta (1995), 1261: 432-434). This fragment contains a well defined bipartite nuclear localisation signal (KRKYSAAKTKVEKKKKKE) and meets the consensus of a nucleolus localisation signal. We have also found inserts that are new sequences which do not have a match in the databases. This is the case of clone 09/07#18 (Figure 5C, D). These cell contain GFP that is tightly localised to the endoplasmic riticulum (ER), as shown by counterstaining with an antibody against the ER marker $\alpha$-calnexin (not shown) (Cannon et al., J. Biol. Chem. (1999), 274: 7537-7544). The insert from this cell line encodes a peptide, 35 amino acids long. It does not have a match in the sequence databases, but contains a predicted trans-membrane motif (PMSIFIQLIYFLLFLFLGVIC) that may occur for the ER specific retention shown by the fusion protein (Dotti et al., Cell (1990), 62: 63-72.

**Example 4**

Construction of the GET vector (GET#1). Using primers A (CATGTTGGCGGCCGCGGTACCGTCGA) and B

(GCCCGGGCGTGAGCAAGGGCGAG) we modified pEGFP-N1 (Clontech) by PCR to introduce a SrfI site between nucleotides three and four of the GFP coding sequence. This insertion shifts the initial ATG codon of the GFP out of frame with the rest of the coding sequence. This ensures that only insert-containing plasmids will express GFP. PCR was carried out with the Expand High Fidelity PCR System (Boehringer). Oligo A also introduced a Not1 site 10 nucleotides upstream of the GFP CDS. The PCR product was purified using the PCR Quiaquick Purification System (Quiagen), ligated with Rapid Ligation Kit (Boehringer) and used to transform Epicurian Coli XL1-Blue (Stratagene), by heat-shock. Transformed cells were plated out in LB-Agar supplemented with 30 μg/ml kanamycin and incubated for 16 hours at 37°C. The modified vector was then isolated by minipreps and the Not1 fragment subcloned into a pQE31 vector previously modified to introduce a Not1 site between the BamHI and KpnI sites with an adaptor made with oligos Not1-1b (GATCGCGGCCGCGTAC) and Not1-8 (GCGGCCGC). The resulting colonies were checked under a transiluminator to test the expression of GFP and the Not1 fragment was then isolated from one of the colonies and subcloned into pEGFP-N1-Not, a modified version of pEGFP-N1 that carries an additional Not1 site inserted in position 635-642.

**Example 5**

Construction of the MmcDNA-GET library. The cDNA was obtained from NIH/3T3 cells by random priming, purified with QIAEX II Gel Extraction Kit, cloned into the SrfI site of the GET#1 vector using the Rapid Ligation Kit (Boehringer) and transformed into E. coli XL1-Blue MR (Stratagene). Plating out a small aliquote of these cells we estimated that the library contained about 420.000 clones of which 1.6% had no insert. The complete library was then plated out onto a sterile Nylon filter laid out on a 24x24 cm plate containing LB supplemented with kanamycin, incubated at 30°C for 24 hours, replicated into another filter and reincubated for 4 hours at 37°C.

The DNA from the library was then purified with Plasmid Maxi Kit (QIAGEN).

**Example 6**

Transfection of HEK293 cells with GET libraries and cloning of cells displaying localised GFP. HEK293 cells were transfected with the GET libraries following the method described by Chen and Okayama (Mol Cell Biol (1987), 7: 2745-2752). The cells were observed twelve to sixteen hours after transfection to check for localised GFP using an inverted LEICA DMI-RBE microscope using a long distance 63x Fluotar objective. The position of the cells of interest was labelled with a diamond pen and then cloned by a combination of manual cloning and serial dilutions, as described in Harlow and Lane (Antibodies: a laboratory manual (1988), Cold Spring Harbour Laboratory Press, N.Y.). In some cases, the cells were first cloned using a fluorescence-activated cell sorter (FACS) and the resulting clones were later analysed to determine the presence of localised GFP.

**Example 7**

Cloning of the DNA fragments encoding subcellular localisation sequences. These were isolated from cloned cells by RT-nested PCR using oligos Fir (AGCTTCGAATTCGCGGCCGCCAACATG) Sec (TATGATCTAGAGTCGCGGCCGCTTTAC) Thi (TAGCGCTACCGGACTCAGATCTCGAGC) and Fou (AAAACCTCTACAAATGTGGTATGGCTG) which flank the SrfI site of the GET#1 vector. mRNA isolation was carried out using the mRNA Capture Kit (Boehringer). The reverse transcriptase reaction and the first round of PCR were carried out using the Titan One Tube RT-PCR Kit with the Expand High

Fidelity PCR System (Boehringer). Oligo Fou was used to prime the RT reaction. The first and second rounds of PCR used oligos Thi and Fou and Fir and Sec as primers. The PCR product was run on an agarose gel and isolated with QIAEX II Gel Extraction Kit (QIAGEN). The isolated fragment was then digested with Not 1 and subcloned into the GET#1 vector to check that the isolated fragment drives the GFP to the expected localisation and for sequencing.

## Claims

1. Process for the detection, cloning and/or sequencing of polypeptides or parts thereof, which drive the subcellular localization of a protein containing such polypeptide or part thereof,
**characterized** in that the process comprises the following steps:
(a) constructing an expression library of random nucleic acids ligated to a reporter gene and contained in a vector molecule,
(b) transfecting a plurality of host cells with the library,
(c) screening for the subcellular localization of the expression product of the nucleic acid in the host cells via detection of a signal produced by the reporter gene,
(d) cloning such cells where the reporter gene signal is detected in a certain subcellular localization, and
(e) cloning and optionally sequencing the nucleic acid insert which encodes the polypeptide or part thereof.

2. Process according to claim 1,
**characterized** in that a cDNA or cDNA fragments are used as random nucleic acids.

3. Process according to claim 1 or 2,
**characterized** in that a eukaryotic or a yeast library is used.

4. Process according to anyone of claims 1 to 3,
**characterized** in that a homologous system of library and cells for the transfection is used.

5. Process according to anyone of claims 1 to 3,
**characterized** in that a heterologous system of library and cells for the transfection is used.

6. Process according to claim 5,
**characterized** in that a Drosophila library is used to transfect mammalian or yeast cells.

7. Process according to anyone of claims 1 to 6,
**characterized** in that a reporter gene leading to a visually detectable signal upon expression is used.

8. Process according to claim 7,
**characterized** in that nucleic acids coding for GFP, BFP, luciferase or YFP are used as reporter gene.

9. Process according to anyone of claims 1 to 8,
**characterized** in that the vector contains an inducible promoter driving the expression of random nucleic acid and marker gene.

10. Process for the identification and/or production of a protein that is localized in a given subcellular localization,
**characterized** in that a nucleic acid coding for a polypeptide or part thereof driving the localization in said given subcellular localization is cloned according to claims 1 to 9 and the nucleic acid is used to detect DNA sequences coding for a protein containing such polypeptide or part thereof.

11. Process according to claim 10,
**characterized** in that for the production of the protein the nucleic acid is expressed in an expression system.

12. Process for directing the subcellular localization of a nucleic acid expression product,
**characterized** in that a polypeptide driving the localization of a protein containing such polypeptide or part thereof is detected, its nucleic

acid sequence is obtained by a process according to anyone of claims 1 to 8, the nucleic acid coding for the polypeptide or part thereof is fused to a nucleic acid coding for a protein to be expressed, and the fusion product is expressed.

13. Process according to claim 12,
**characterized** in that a nucleic acid coding for the polypeptide or part thereof and a reporter gene is fused to the nucleic acid coding for a protein to be expressed.

14. Process according to claim 12,
**characterized** in that a reporter gene the expression product of which is visually detectable is used.

15. Process according to anyone of claims 12 to 14,
**characterized** in that the fusion product contains a proteolytic cleavage site between the protein to be expressed and the polypeptide or part thereof and/or reporter gene product.

16. Vector for the expression of a desired protein wherein the vector contains a specific site into which a DNA encoding said desired protein can be inserted,
**characterized in that** the vector further comprises a DNA sequence encoding a polypeptide or a part thereof which drives the subcellular localization of a protein containing such polypeptide or part thereof, which DNA sequence is positioned in such a way that a fusion protein of desired protein and polypeptide or part thereof is encoded.

17. Vector according to claim 16,
**characterized in that** the vector is a eucaryotic vector.

18. Vector according to claim 16 or 17,

**characterized in that** the vector further comprises a reporter gene positioned in such a way that a fusion protein of desired protein and polypeptide or part thereof and reporter gene product is encoded.

19. Vector according to claim 18,
**characterized in that** the reporter gene product is visually detectable.

20. Vector according to anyoner of claims 16 to 19,
**characterized in that** the vector further contains sequences encoding proteolytic cleavage sites between one or more of the constituents of the fusion protein.

21. Cell line,
**characterized in that** it is transfected with a vector according to anyone of claims 16 to 20, encoding a fusion protein of at least a polypeptide or part thereof driving the localisation to a given subcellular localisation and a desired protein.

22. Kit for the expression of a desired protein in a desired localisation of a host cell,
**characterized in that** it contains a vector according to anyone of claims 16 to 20 or a cell line according to claim 21 optionally together with other components and/or buffers for the protein expression.

23. Collection of cell lines according to claim 21.

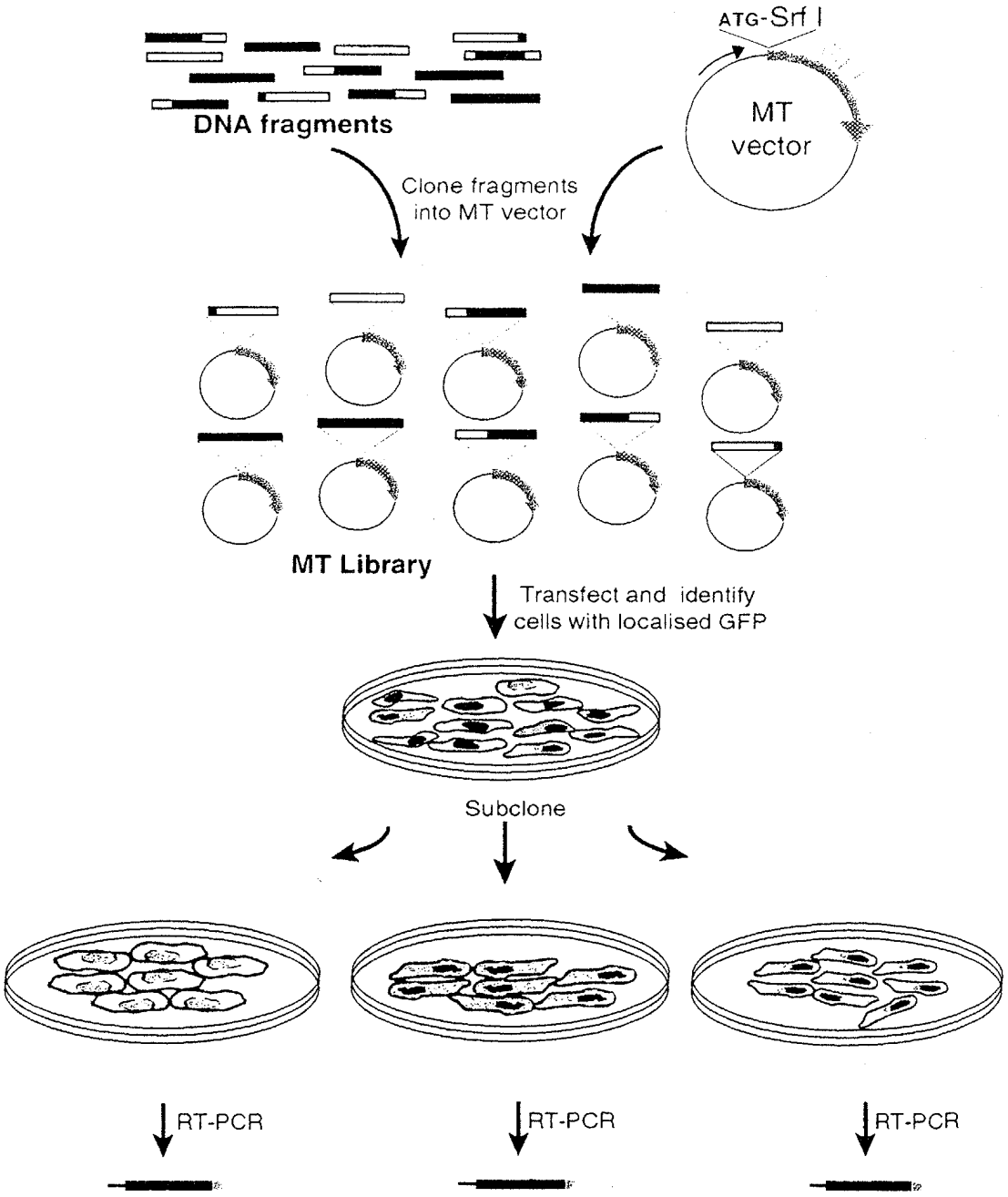

ATG-Srf I
DNA fragments
MT vector
Clone fragments into MT vector
MT Library
Transfect and identify cells with localised GFP
Subclone
RT-PCR
RT-PCR
RT-PCR
Isolated DNA fragments encoding subcellular localisation signals

Figure 1

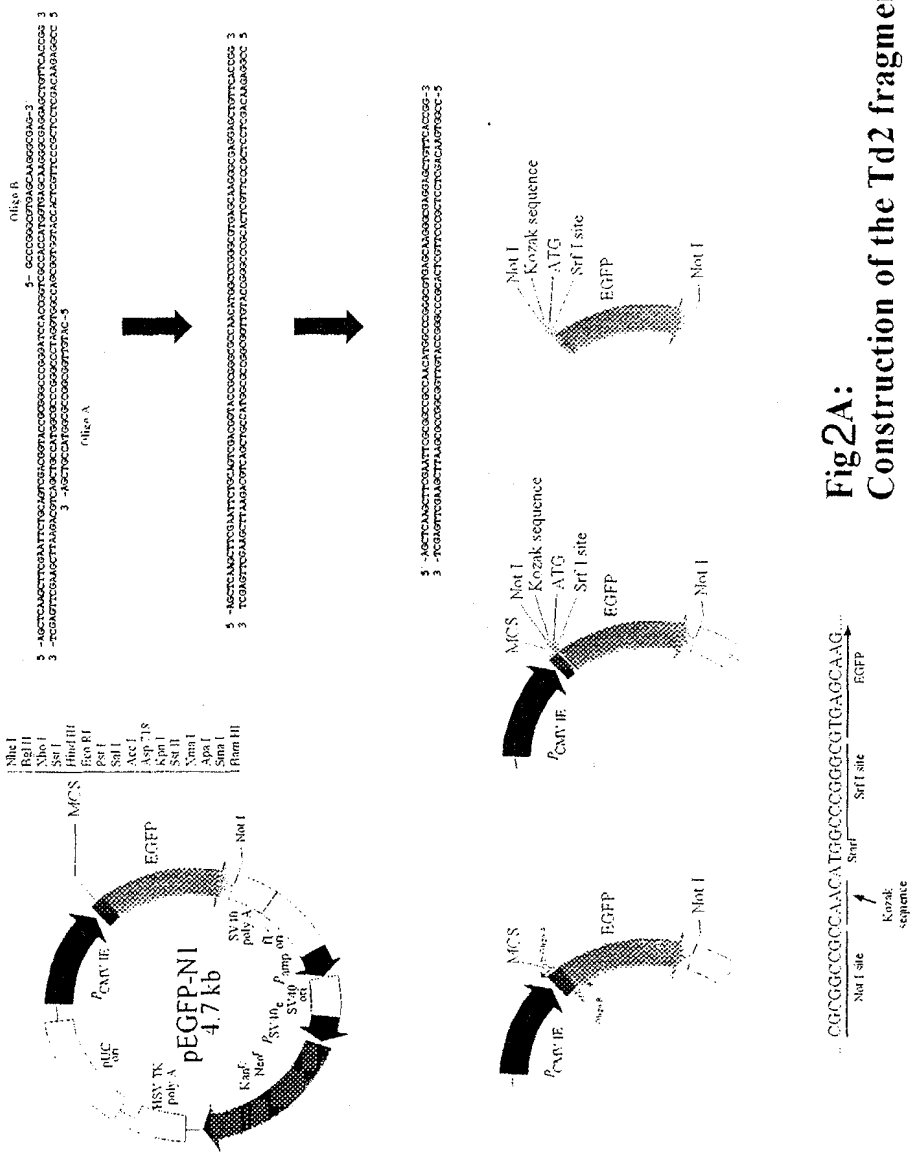



**Fig2A: Construction of the Td2 fragment**

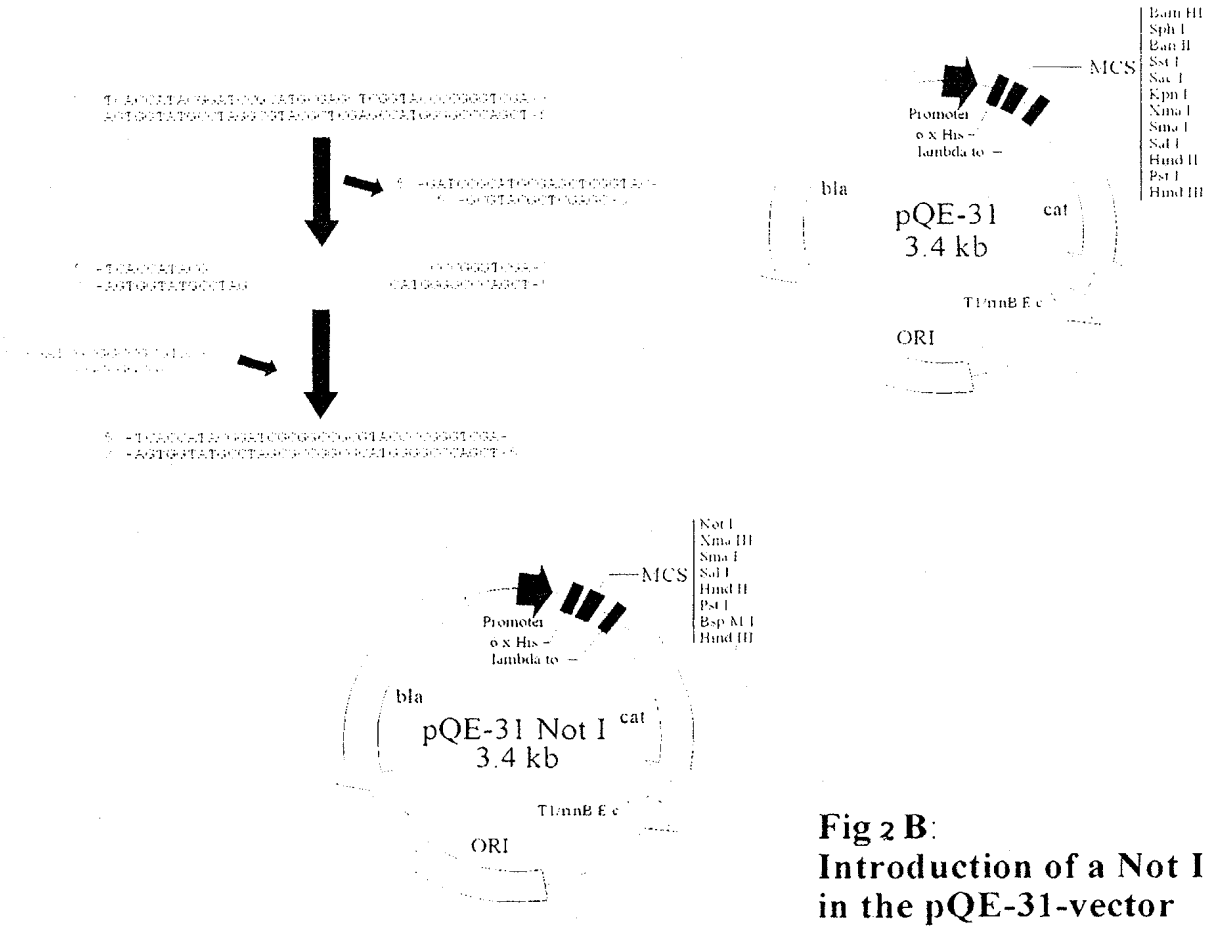



**Fig 2 B:**
**Introduction of a Not I site in the pQE-31-vector**

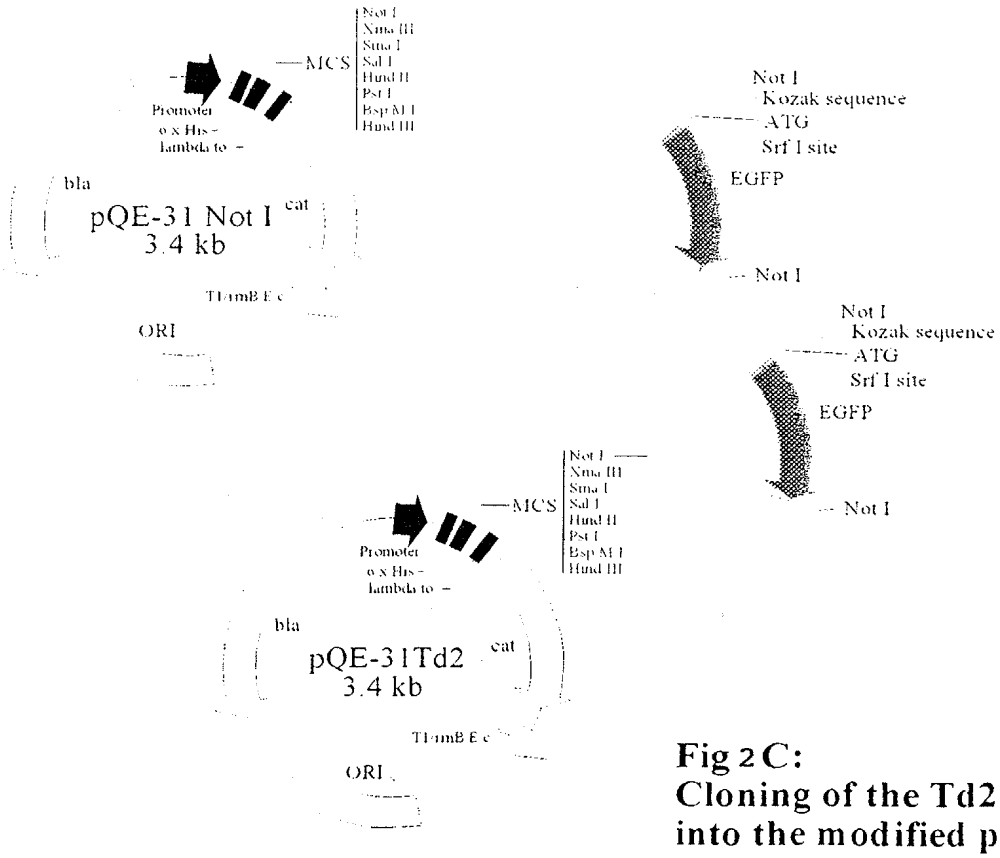



**Fig 2C:**
**Cloning of the Td2 fragment into the modified pQE-31 vector**

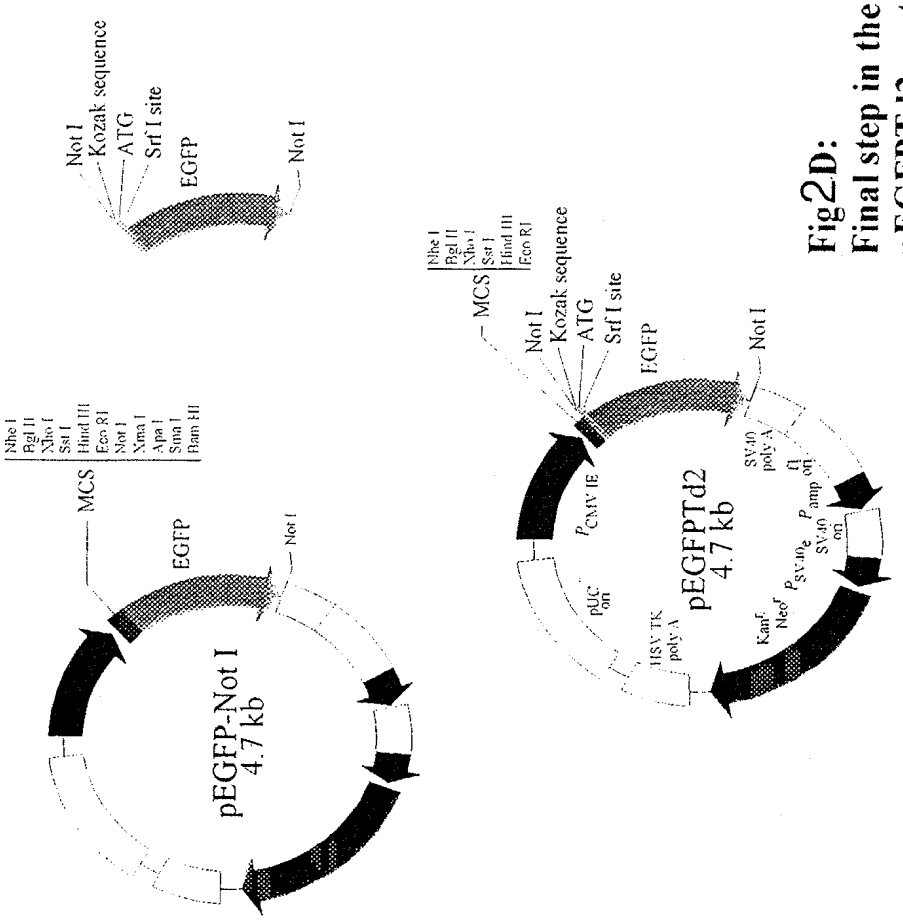



**Fig2D:**
**Final step in the production of the pEGFPTd2 vector**

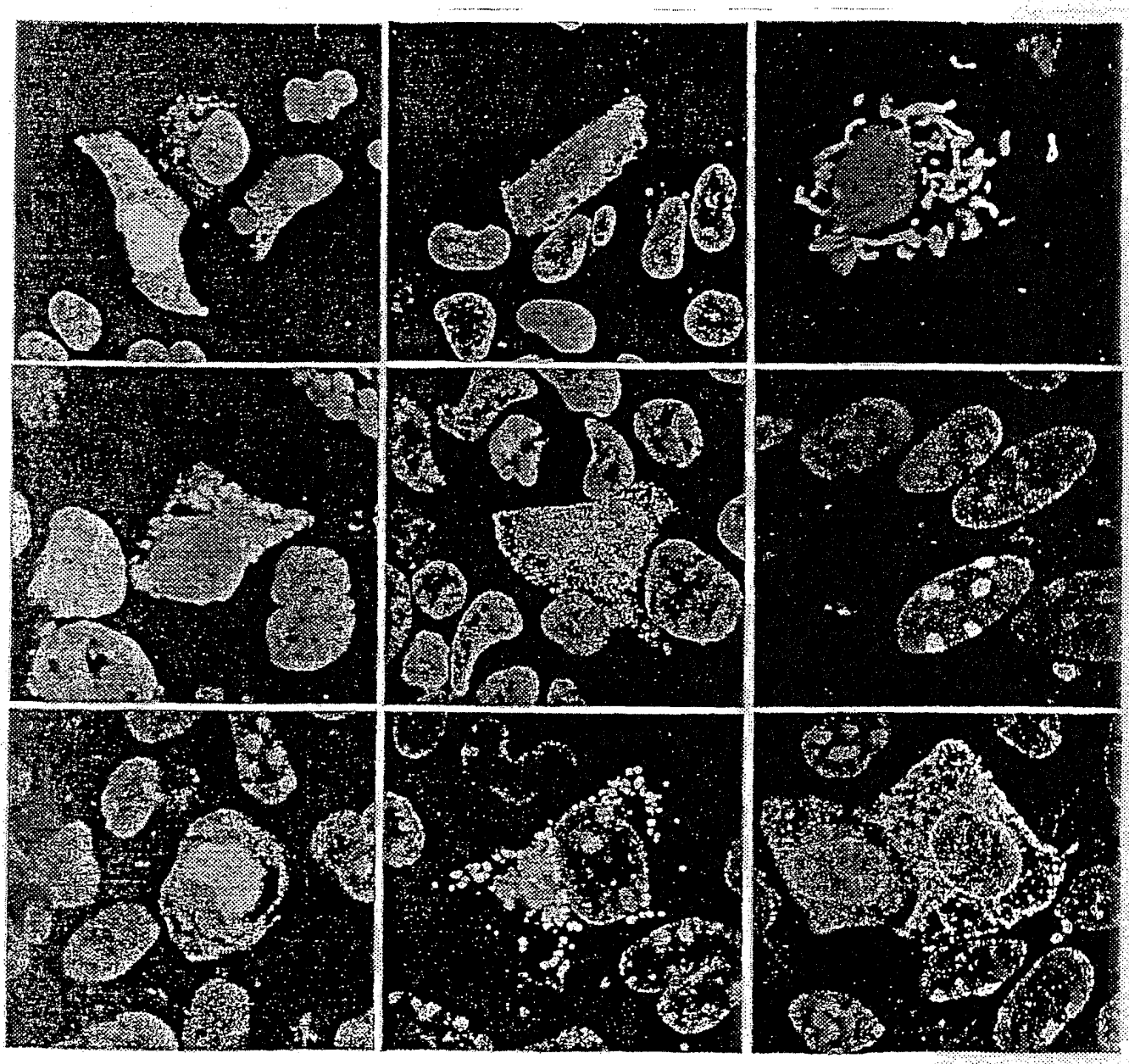


Figure 3

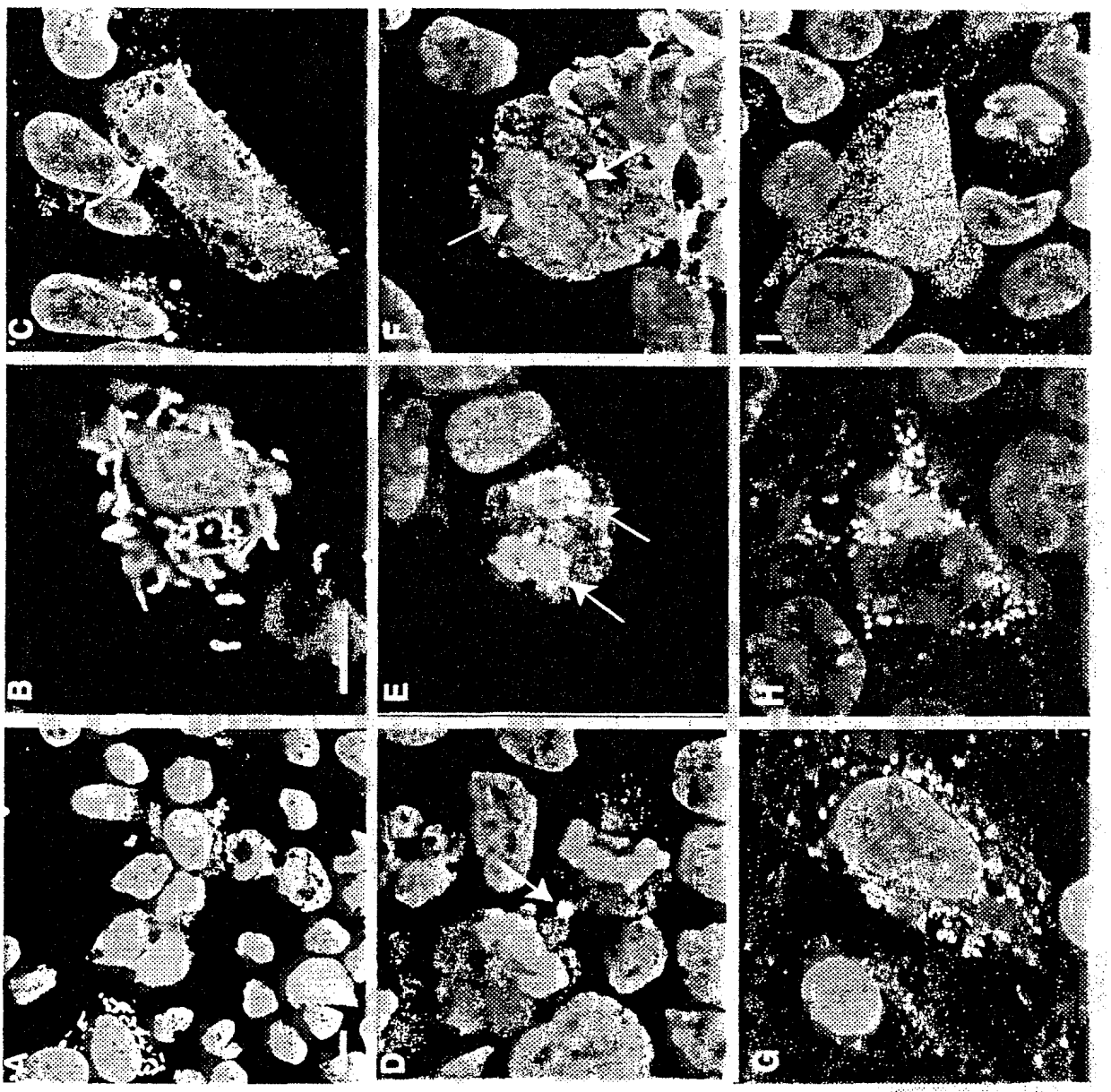
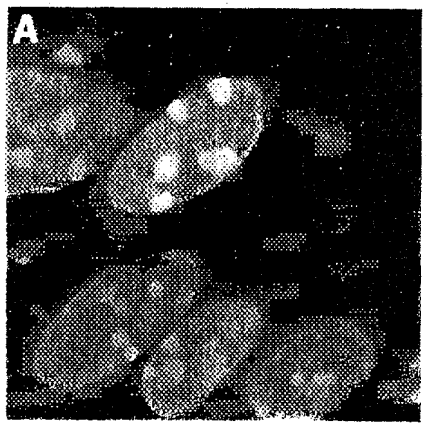
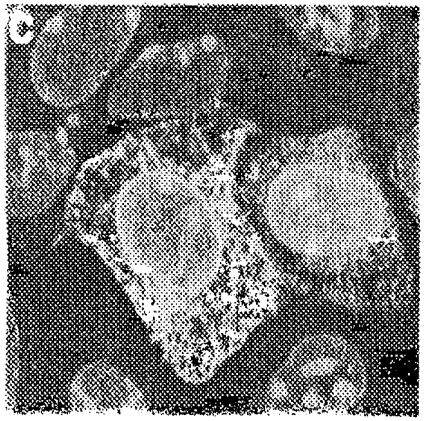



Figure 4




B

1 RKALY KV 25
26 LATVTKTVGGDKNGGTRVVKLRKMP 50
51 RYYPTEDVPRKLLSHGKKP 69




D

1 MA RC 25
26 SCIKENIRGV 35

Figure 5

# INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 00/02607

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/10 C12N15/12 C12N15/62 C07K14/435

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | SAWIN K E AND NURSE P: "Identification of fission yeast nuclear markers using random polypeptide fusions with green fluorescent protein" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 94, no. 26, 24 December 1996 (1996-12-24), pages 15146-15451, XP002121720 the whole document<br>---<br>-/-- | 1-4, 7-14, 16-19, 21-23 |

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 10 July 2000 | 18/07/2000 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL – 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,<br>Fax: (+31-70) 340-3016 | Oderwald, H |

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | CHA H J ET AL.: "Purification of human interleukin-2 fusion protein produced in insect larvae is facilitated by fusion with green fluorescent protein and metal affinity ligand" BIOTECHNOLOGY PROGRESS, vol. 15, no. 2, 19 March 1999 (1999-03-19), pages 283-286, XP002121721 ISSN: 8756-7938 the whole document | 1-5,7-23 |
| A | GERDES H -H ET AL: "GREEN FLUORESCENT PROTEIN: APPLICATIONS IN CELL BIOLOGY" FEBS LETTERS, vol. 389, 1 January 1996 (1996-01-01), pages 44-47, XP000770275 ISSN: 0014-5793 the whole document | |
| P,X | BEJARANO L A AND GONZALEZ C: "Motif Trap: A rapid method to clone motifs that can target proteins to defined subcellular localisations." JOURNAL OF CELL SCIENCE, vol. 112, no. 23, December 1999 (1999-12), pages 4207-4211, XP000923119 ISSN: 0021-9533 the whole document | 1-23 |

# INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 00/02607

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/10 C12N15/12 C12N15/62 C07K14/435

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | SAWIN K E AND NURSE P: "Identification of fission yeast nuclear markers using random polypeptide fusions with green fluorescent protein" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 94, no. 26, 24 December 1996 (1996-12-24), pages 15146-15451, XP002121720 the whole document<br>---<br>-/-- | 1-4, 7-14, 16-19, 21-23 |

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 10 July 2000 | 18/07/2000 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL – 2280 HV Rijswijk<br>Tel. (+31–70) 340–2040, Tx. 31 651 epo nl,<br>Fax: (+31–70) 340–3016 | Oderwald, H |

3

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication,where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | CHA H J ET AL.: "Purification of human interleukin-2 fusion protein produced in insect larvae is facilitated by fusion with green fluorescent protein and metal affinity ligand" BIOTECHNOLOGY PROGRESS, vol. 15, no. 2, 19 March 1999 (1999-03-19), pages 283-286, XP002121721 ISSN: 8756-7938 the whole document | 1-5,7-23 |
| A | GERDES H -H ET AL: "GREEN FLUORESCENT PROTEIN: APPLICATIONS IN CELL BIOLOGY" FEBS LETTERS, vol. 389, 1 January 1996 (1996-01-01), pages 44-47, XP000770275 ISSN: 0014-5793 the whole document | |
| P,X | BEJARANO L A AND GONZALEZ C: "Motif Trap: A rapid method to clone motifs that can target proteins to defined subcellular localisations." JOURNAL OF CELL SCIENCE, vol. 112, no. 23, December 1999 (1999-12), pages 4207-4211, XP000923119 ISSN: 0021-9533 the whole document | 1-23 |

PATENT COOPERATION TREATY

# PCT

## INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

| Applicant's or agent's file reference | **FOR FURTHER ACTION** | see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below. |
|---|---|---|
| 19741P WO | | |
| International application No. | International filing date *(day/month/year)* | (Earliest) Priority Date *(day/month/year)* |
| PCT/EP 00/ 02607 | 23/03/2000 | 23/03/1999 |

Applicant

EUROPÄISCHES LABORATORIUM FÜR MOLEKULARBIOLOGIE..

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 3 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

a. With regard to the **language,** the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

[ ] the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

[X] contained in the international application in written form.

[ ] filed together with the international application in computer readable form.

[ ] *furnished subsequently to this Authority in written form.*

[X] furnished subsequently to this Authority in computer readble form.

[X] the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

[X] the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. [ ] **Certain claims were found unsearchable** (See Box I).

3. [ ] **Unity of invention is lacking** (see Box II).

4. With regard to the **title,**

[X] the text is approved as submitted by the applicant.

[ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract,**

[X] the text is approved as submitted by the applicant.

[ ] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No. 1

[ ] as suggested by the applicant.

[X] because the applicant failed to suggest a figure.

[ ] because this figure better characterizes the invention.

[ ] None of the figures.

# INTERNATIONAL SEARCH REPORT

International Application No
P[redacted]P 00/02607

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/10 C12N15/12 C12N15/62 C07K14/435

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | SAWIN K E AND NURSE P: "Identification of fission yeast nuclear markers using random polypeptide fusions with green fluorescent protein" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 94, no. 26, 24 December 1996 (1996-12-24), pages 15146-15451, XP002121720 the whole document<br>---<br>-/-- | 1-4, 7-14, 16-19, 21-23 |

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 10 July 2000 | 18/07/2000 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL – 2280 HV Rijswijk<br>Tel. (+31–70) 340–2040, Tx. 31 651 epo nl,<br>Fax: (+31–70) 340–3016 | Oderwald, H |

3

Form PCT/ISA/210 (second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

International Application No
P[redacted]P 00/02607

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication,where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | CHA H J ET AL.: "Purification of human interleukin-2 fusion protein produced in insect larvae is facilitated by fusion with green fluorescent protein and metal affinity ligand" BIOTECHNOLOGY PROGRESS, vol. 15, no. 2, 19 March 1999 (1999-03-19), pages 283-286, XP002121721 ISSN: 8756-7938 the whole document | 1-5,7-23 |
| A | GERDES H -H ET AL: "GREEN FLUORESCENT PROTEIN: APPLICATIONS IN CELL BIOLOGY" FEBS LETTERS, vol. 389, 1 January 1996 (1996-01-01), pages 44-47, XP000770275 ISSN: 0014-5793 the whole document | |
| P,X | BEJARANO L A AND GONZALEZ C: "Motif Trap: A rapid method to clone motifs that can target proteins to defined subcellular localisations." JOURNAL OF CELL SCIENCE, vol. 112, no. 23, December 1999 (1999-12), pages 4207-4211, XP000923119 ISSN: 0021-9533 the whole document | 1-23 |